Total observed error, total allowable error, and QC rules for canine serum and urine cortisol achievable with the Immulite 2000 Xpi cortisol immunoassay

Jeremie Korchia,[1] Kathleen P. Freeman

Abstract. Determining a simple quality control (QC) rule for daily performance monitoring depends on the desired total allowable error (TEa) for the measurand. When no consensus TEa exists, the classical approach of QC rule validation cannot be used. Using the results of previous canine serum and urine cortisol validation studies on the Immulite 2000 Xpi, we applied a reverse engineering approach to QC rule determination, arbitrarily imposing sigma = 5, and determining the resulting TEa for the QC material (QCM; TEa_{QCM}) and the resulting probability of error detection (P_{ed}) for each QC rule. For the simple QC rule $1_{2.5S}$ with $P_{ed} = 0.96$ and probability of false rejection (P_{fr}) = 0.03, the associated TEa_{QCM} were 20% and 35% for serum and 28% and 24% for urine QCM1 and QCM2. If these levels of TEa_{QCM} are acceptable for interpretation of patient sample results, then users can internally validate the $1_{2.5S}$ QC rule, provided that their QCM CVs and biases are similar to ours. Otherwise, more stringent QC rules can be validated by using a lower sigma to lower the TEa_{QCM}. With spiked samples (relevant cortisol concentrations in the veterinary patient matrix) at 38.6 and 552 nmol/L of cortisol, TEa_{QCM} at sigma = 5 were much higher (54% and 40% for serum; 90.3% and 42.8% for urine). Spiked samples generate TEa that is probably too high to be suitable for daily QC monitoring; however, it is crucial to verify spiked sample observed total error (TEo; 26% and 18% for serum, 60% and 30% for urine) < TEa_{QCM}, and to use spiked sample TEo for patient result interpretation. In the absence of consensus TEa for cortisol in dogs, we suggest the use of a $1_{2.5S}$ rule, provided that users accept the associated level of TEa_{QCM} also as clinical TEa for results interpretation.

Keywords: Cushing; cortisol; dogs; endocrinology; quality control; total allowable error.

Quantitative medical laboratory test results are best interpreted in light of the laboratory test performance (observed total error, TEo), or in light of the minimal required laboratory test performance (total allowable error, TEa) when the latter is available.[17] In human medicine, consensus TEa has been established for many measurands.[16,25] In veterinary medicine, consensus TEa have been determined relatively recently for hematology[11] and biochemistry.[8] However, there is, to date, no consensus TEa recommendations for veterinary endocrinology.

The concept of TEa is defined in the American Society for Veterinary Clinical Pathology (ASVCP) guidelines[1] as "a quality goal that sets a limit for combined imprecision (random error) and bias (systematic error) that is tolerable in a single measurement to ensure clinical usefulness." In contrast, TEo is commonly calculated from imprecision (SD) and inaccuracy (bias) by the formula: TEo = 2SD + absolute bias, in which all elements are expressed in quantitative units of the assessed measurand (Fig. 1). When normalized by the mean of the target value, SD becomes the coefficient of variation (CV), and the formula becomes: TEo(%) = 2CV(%) + absolute bias(%).[17] Thus, TEa corresponds with a consensus about the level of allowed

combined errors to assure sufficient quality to achieve a useful clinical interpretation. The recommended TEa also is typically used as the TEa for the quality control material (QCM; TEa_{QCM}) to generate acceptable QC rules based on the QCM precision and bias. The normal distribution (assumed to be present with measurement replicates) results in 95% of the measurements occurring within an interval corresponding to the biased result ±2CV.[17]

TEa and TEo are related through the sigma metric, a unitless metric that can be applied in quality management. The sigma metric, or just sigma (σ), is defined by the formula: σ = (TEa – absolute bias)/CV. In other words, sigma quantifies the "room" left for CVs once the bias has been removed from TEa (Fig. 2A). The higher the sigma (indicating high testing system capability), the easier the monitoring of

Texas A&M Veterinary Medical Diagnostic Laboratory, Texas A&M University, College Station, TX, USA (Korchia); SYNLAB-VPG/Exeter, Exeter, United Kingdom (Freeman).

[1]Corresponding author: Jeremie Korchia, Department of Microbiology, Immunology and Pathology, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, CO 80521, USA. jeremie.korchia@yahoo.fr
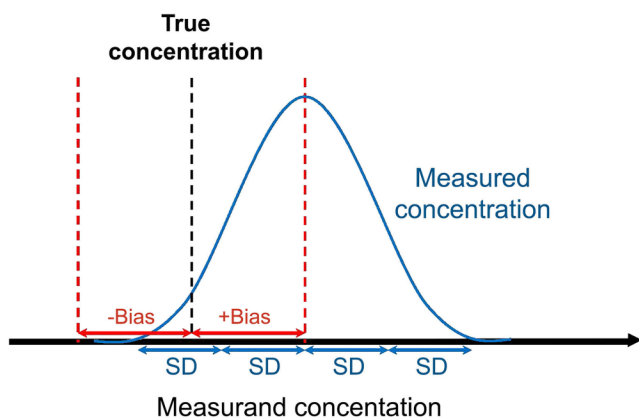
Figure 1. Origin of the observed total error (TEo) formula. TEo is calculated from imprecision (SD) and inaccuracy (bias) by the formula: TEo=2SD+absolute bias, in which all elements are expressed in units of the assessed measurand. When normalized by the mean of the target value, SD becomes the coefficient of variation (CV), and the formula becomes: TEo(%)=2CV(%)+absolute bias(%).[17] 4SD includes 95% of the results (mean±2SD) when values are normally distributed, which is assumed for measured replicates.

Figure 2. A. Classical approach and **B.** reverse approach for QC rule validation. EZ Rules 3 QC design software; OPSpec charts=operational process specification charts; Ped=probability of error detection; Pfr=probability of false rejection; QC=quality control; TEa=total allowable error; TEo=observed total error; σ=sigma metric.

performance quality.[21,22] A higher sigma translates to the ability to use less stringent QC rules to achieve satisfactory probability of error detection (P_{ed}) and probability of false rejection (P_{fr}).[24] The sigma formula can be manipulated, or reversed, to solve for TEa as: TEa=absolute bias+σCV (Fig. 2B). Written as such, the similarity with TEo=|bias|+2CV is obvious. The link between TEo and TEa through sigma is then: TEo=TEa when sigma=2. For a testing system to be considered reliable for commercial use, sigma of the system needs to be at least 3; it often lies between 3 and 5.[19] Sigma of 5 is by definition good or excellent; sigma of 6 is defined as "world class" quality.

For definition purposes in our manuscript, the testing system (referred to as the system) corresponds with the sum of elements allowing measurements to be performed, including the analyzer and the reagents, as well as the analyzer maintenance and the skills of the technician. The performance of the system can be tested by diverse samples (patient, commercial QCM). In the case of a QCM sample, strictly speaking we should talk about "CV and bias of the QCM with a given system" (the same QCM has different CV and bias depending on the system with which measurements are performed). For language simplification purposes, we will talk of CV, bias, and TE of a sample instead.

According to ASVCP guidelines,[1] TEa should be determined as a quality goal prior to assessment of a method, and can be derived mathematically from clinical decision limits or interpretation thresholds (IT). A method is judged acceptable when TEo<TEa. Admittedly, consensus recommendations for TEa are not available for cortisol in serum and in urine in dogs (not addressed in the last ACVIM consensus about Cushing disease in dogs[3]); however, the effect on the clinical
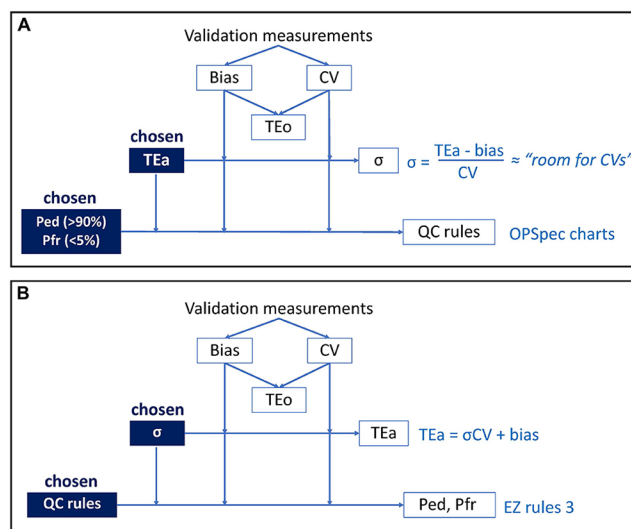
outcome is not the only model available for analytical performance specifications. A consensus statement[13] from human medicine reports 3 models for analytical performance specifications, namely model 1 "based on the effect of analytical performance on clinical outcomes," model 2 "based on components of biological variation of the measurand," and model 3 "based on state-of-the-art" or "highest level of analytical performance technically achievable." The consensus statement specifies that preference should be given to models 1 and 2, and that, in some situations, it can be advantageous to combine the models. For cortisol in dogs, given that a consensus TEa has not been defined, model 1 cannot be used. For serum cortisol in dogs, the concentration is variable (rhythms, stress, etc.), and most importantly, it is assessed via dynamic testing, precluding the use of model 2. Following the hierarchy of the consensus statement, we aimed to use model 3 focusing on state-of-the-art in an attempt to characterize analytical performance specifications. Of note, the consensus statement emphasizes that "the hierarchy assumes that high quality studies or data are available for each model," and that "the quality of the available evidence behind each model may modulate the selection of the best approach."[13]

For patient samples, characterizing TEo (analytical performance technically achievable) at relevant IT[4] is a good starting point: it is easier to start determining if TEa should be higher or lower than a given TEo, rather than to determine TEa arbitrarily. Thus, if the method can be used clinically, TEo must be small enough relative to TEa such that sigma is >2. In other words, TEo is the lower end of the range in which to choose a suitable TEa for an acceptable method.

The same is true for QCM samples. Additionally, the TEa_{QCM} can be determined despite the absence of a consensus

TEa for the measurand of interest, based on the QC rule used and level of P_{ed} desired. When consensus TEa exists, $TEa_{QCM} = TEa$, and it is the TEa_{QCM} that drives the selection of the candidate QC rule based on satisfactory P_{ed}. When consensus TEa does not exist, in theory, QC rules cannot be validated; in practice, veterinary laboratories often arbitrarily choose to use a simple QC rule, usually 1_{2s}, $1_{2.5s}$, or 1_{3s}. This is a laudable effort; however, any arbitrarily chosen rule will necessarily operate with unknown P_{ed}, which reflects the efficacy of the QC rule. Given that characterizing P_{ed} requires TEa_{QCM} and the latter is unknown, laboratories usually ignore both of them. Yet, if all of the other parameters are fixed to the state-of-the-art model, it becomes possible to calculate TEa_{QCM} or P_{ed} as the only unknown entity in the equation. This is a "reverse approach" to candidate QC rule determination; unlike the classical approach, it does not validate the QC rule per se, however it extensively characterizes the efficacy of the used QC rules or the TEa_{QCM} resulting from the system used in the state-of the art condition, offering users the possibility of subjectively selecting a QC rule with high enough P_{ed} or low enough TEa_{QCM} for their purposes (i.e., internal validation).

Based on the characterization of sigma as good-to-excellent,[21,22] our model uses TEa_{QCM} yielding a sigma of 5 as the state-of-the-art performance, and uses the resulting P_{ed} of each QC rule to determine candidate QC rule acceptance. This model allows comparison of sigma-dependent TEa_{QCM} with our expert opinion of what TEa could be for canine cortisol, as well as comparison with state-of-the-art performance of the method. We used the databases of 2 of our previous studies validating cortisol measurement in canine serum[9] and in canine urine[10] with the Immulite 2000 Xpi (Siemens), investigating precision, bias, and TEo across the cortisol reportable range on spiked matrices, and at 2 levels of QCM. We aimed to:

- Use the reverse approach to validate QC rules for QC monitoring with commercial QCMs.
- Illustrate the difference between state-of-the-art performance (precision, bias) of commercial QCMs and spiked samples at relevant cortisol concentrations within real animal matrices.
- Discuss levels of TEa based on state-of-the art performance of the method.
- Use the example of cortisol in dogs to illustrate the complex relationships between the concepts involved in statistical QC monitoring to promote understanding within the veterinary community.

Materials and methods

Study overview

We used data for canine serum and urine cortisol determined previously in validation studies with the Immulite 2000 Xpi.[9,10] Briefly, both studies consisted of 3 phases (Fig. 3): 1) a spiking-recovery phase on "cortisol-free" matrices (pooled canine sera from ACTH stimulation tests [ACTHST] and from low-dose dexamethasone suppression tests [LDDST] all with undetectable cortisol, and urine from an Addisonian dog with undetectable cortisol) from which we investigated reportable range and linearity, precision, spiking-recovery bias (SR), and detection limit (the calibration verifier mode of the Immulite 2000 Xpi was used to measure cortisol concentration outside of the manufacturer's reportable range); 2) an interlaboratory comparison phase, from which we investigated the average bias (AB) between institutions, as well as the range-based bias (RB), meaning the biases observed in multiple limited concentration ranges; and 3) a QCM phase with commercial QCMs and spiked patient sample QCMs, each with 2 levels for each medium. Commercial QCMs consisted of K9CON (Immulite Systems Control; Siemens) for serum, and Liquicheck (Bio-Rad) for urine. Importantly, the K9CON had target values for the Immulite 2000, hence the QCM bias could be determined, in addition to the between-run precision. On the other hand, there is currently no QCM with target values for urine cortisol on an Immulite analyzer. The QCM for urine cortisol had target values for several different measurement methods but not for the Immulite 2000 Xpi, hence the QCM bias could not be determined. Therefore, to calculate the TEo for QCM (TEo_{QCM}) for urine, we compared the 3 biases from the 2 previous phases (SR bias, AB, RB), and elected to use the AB.[10] We spiked cortisol-free canine matrices (serum and urine) to produce patient sample QCMs at the common IT concentrations of serum cortisol. Importantly, these artificial, spiked samples were identical to true canine patient serum and urine samples at these cortisol concentrations.

We used the CV,[18] bias,[18] and TEo results for both QCM levels in both media to determine optimized TEa_{QCM} of the system at $\sigma = 5$, and used the software EZ Rules 3 (Westgard QC) to determine the P_{ed} of each of 4 QC rules to determine which was the best candidate. The P_{ff} is a function of QC rules and of the numbers of QCM levels, independent of the method performance (for 2 QCM levels: 0.09 for 1_{2S}, 0.03 for $1_{2.5S}$, 0 for 1_{3S}, 0.01 for $1_{3S}/2_{2S}/R_{4S}$). We also investigated similarly spiked samples at 2 cortisol concentrations, 38.6 nmol/L (1.4 µg/dL) and 552 nmol/L (20 µg/dL), that we abbreviated L4 and L8, respectively (according to our dilution scheme[9,10]), as potential QCMs in both media.

Immunoassays

Immulite 2000 Xpi cortisol immunoassay (Siemens): chemiluminoassay for cortisol, is a competitive heterogeneous phase assay, using a surface-bound capture anti-cortisol leporine polyclonal antibody and cortisol-alkaline phosphatase as tracer. This assay was validated by the manufacturer for human serum cortisol. It cross-reacts with prednisone (and prednisolone metabolized to prednisone). The cortisol molecule is identical in humans and in dogs,[12] hence the use of this immunoassay in dogs is appropriate. This assay was used for all of the measurements of serum cortisol in both of the
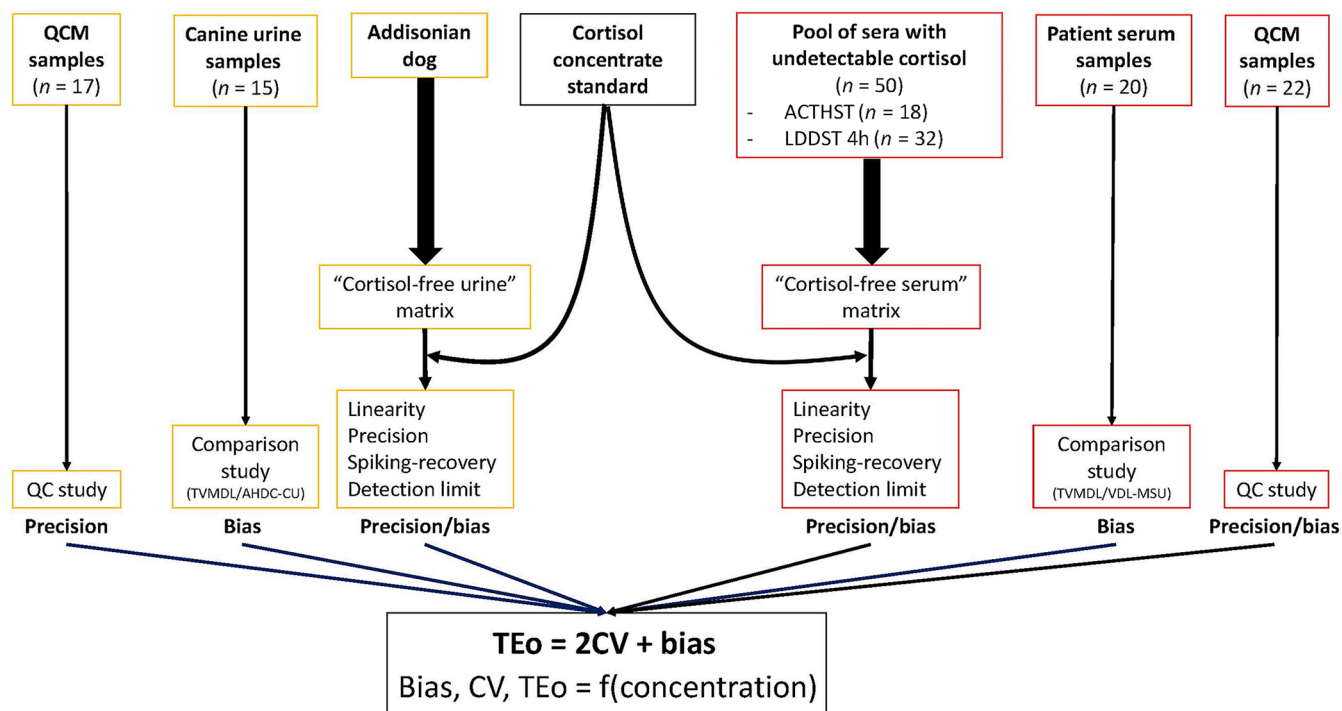
Figure 3. Canine serum cortisol and canine urine cortisol validation study overview. Our study was performed using the databases of 2 previous studies. ACTHST = ACTH stimulation test; AHDC-CU = Animal Health Diagnostic Center of Cornell University; LDDST = low-dose dexamethasone suppression test; QC = quality control; QCM = quality control material; TEo = observed total error; TVMDL = Texas A&M Veterinary Medical Diagnostic Laboratory; VDL-MSU = Veterinary Diagnostic Laboratory of Michigan State University.

Table 1. Types of computed observed total error functions of the between-run CV in different media (rows) and different considered biases (columns).

TEo calculation	Spiking-recovery bias	Average bias	Range-based bias	QCM between-run bias
Between-run CV for serum cortisol	TEo_{SR} (L4, L8)	TEo_{AB} (L4, L8)	TEo_{RB} (equivalent L4, L8*)	TEo_{QCM} (QCM1, QCM2)
Between-run CV for urine cortisol	TEo_{SR} (L4, L8)	TEo_{AB} (L4, L8)	TEo_{RB} (equivalent L4, L8*)	$TEo_{QCM.SR}$ $TEo_{QCM.AB}$ $TEo_{QCM.RB}$ (QCM1, QCM2)†

AB = average bias; L4 = cortisol concentration of 38.6 nmol/L (1.4 µg/dL); L8 = cortisol concentration of 552 nmol/L (20 µg/dL); QCM = quality control material; RB = range-based bias; SR = spiking-recovery bias; TEo = observed total error.
* Ranged-based bias from groups of the comparison study of the closest ranges from those spiked levels.
† There is no target value for urine cortisol QCM with the Immulite 2000 Xpi; thus, to calculate TEo, we used 3 other types of biases: SR and RB at a roughly similar concentrations to those of QCM1 and QCM2, and AB. The latter was elected for QC rule determination for urine cortisol.

laboratories used in our previous reports (Texas A&M Veterinary Medical Diagnostic Laboratory [TVMDL; College Station, TX, USA]; Veterinary Diagnostic Laboratory of Michigan State University [VDL-MSU; East Lansing, MI, USA]), as well as for measurement of urine cortisol at TVMDL.

Immulite 1000 cortisol immunoassay (Siemens): According to the manufacturer's technical information, this assay is similar to the Immulite 2000 cortisol immunoassay, except that kits cannot be used interchangeably. This is the assay used at the reference laboratory, the Animal Health Diagnostic Center of Cornell University (AHDC-CU; Ithaca, NY, USA), in the canine urine cortisol validation study.

Validation databases for serum and urine cortisol in dogs

In this third study, we used results of 2 previous studies (validation on the Immulite 2000 Xpi of canine serum[9] cortisol and canine urine[10] cortisol, respectively) as sources for additional theoretical and mathematical investigation.

CV, biases, and TEo results for canine serum and canine urine cortisol. We provide a summary of the different types of CV and different types of biases[16] that we used to calculate the different types of TEo (Table 1), the results of between-run CV

Table 2. Results from canine serum[9] and canine urine[10] cortisol validation studies, used as materials in our current study.

A. CV (between-run), bias, and TEo results for 2 clinically relevant canine serum cortisol concentrations and 2 QCM levels

Spiked serum samples

Cortisol level	Target values, nmol/L (µg/dL)	Precision (%) Between-run CV	Bias (%) SR	RB	AB	TEo (%) TEo_{SR}	TEo_{RB}	TEo_{AB}
L4	38.6 (1.4)	9.53	−6.50	−13.8	−2.93	25.6	32.8	22.0
L8	552 (20)	7.42	3.08	−4.47		17.9	19.3	17.7

QCM

Cortisol level	Target values,* nmol/L (µg/dL)	Precision (%) Between-run CV	Bias (%)†	TEo_{QCM} (%)
QCM1	193 (7.0)	4.08	−0.27	8.5
QCM2	389 (14.1)	7.01	−0.52	14.5

B. CV (between-run), bias, and TEo results for 2 canine urine cortisol concentrations and 2 QCM levels

Spiked urine samples

Cortisol level	Target values, nmol/L (µg/dL)	Precision (%) Between-run CV	Bias (%) SR	RB	AB	TEo (%) TEo_{SR}	TEo_{RB}	TEo_{AB}
L4	38.6 (1.4)	9.55	42.3	8.67*	−1.09	61.4	27.8	20.2
L8	552 (20)	4.60	19.8	−10.7		28.9	19.9	10.3

QCM

Cortisol level	Target values,* nmol/L (µg/dL)	Precision (%) Between-run CV	Bias (%)† SR	RB	AB	TEo_{QCM} (%) $TEo_{QCM.SR}$	$TEo_{QCM.RB}$	$TEo_{QCM.AB}$
QCM1	193 (7.0)	5.42	35.8	8.67*	−1.09	46.6	19.5	11.9
QCM2	389 (14.1)	4.51	19.8	−10.7		28.7	19.7	10.1

For section A, we used the commercial QCM K9CON (Immulite systems control; Siemens), which has target values available. In serum, TEo_{SR}, TEo_{RB}, and TEo_{AB} are roughly similar; TEo is ~30% for L4 and ~20% for L8.

For section B, we used the commercial QCM Liquicheck (Bio-Rad), for which target values for Immulite analyzers are not available. Means are provided to give an idea of the tested level. Because of the absence of target values, the QCM bias could not be assessed. Instead, TEo was calculated using the biases from the spiking-recovery study (SR) and the comparison study (RB and AB). Unlike in serum,[9] in urine 10 TEo_{SR}, TEo_{RB}, and TEo_{AB} are mismatching because of widely different corresponding biases. The SR bias may be too high and irrelevant in a clinical setting for which comparisons are made between analyzers. The AB may be too low as resulting from the average of opposite biases, and thus irrelevant as neglecting the impact on the measurand concentration on the bias. The RB bias may be the most relevant clinically and would warrant further investigation; of note, the RB bias yields TEo in urine roughly equivalent to serum for L4 (~30%) and L8 (~20%).

AB = average bias; L4 = cortisol concentration of 38.6 nmol/L (1.4 µg/dL); L8 = cortisol concentration of 552 nmol/L (20 µg/dL); RB = range-based bias; SR = spiking-recovery bias; TEo = observed total error.

* Closest available bias.

† There were no target values provided by the manufacturer of the QCM for canine urine cortisol.

and of the 3 types of biases (SR, RB, AB) for both chosen spiked concentrations (L4, L8) and both commercial QCM levels (QCM1, QCM2) in both media (Table 2A), and the corresponding TEo (Table 2B).

QC rule and sigma metric determination by the classical approach of QC rule validation. Candidate QC rules were explored in previous studies[9,10] with a classical approach (Fig. 2A), for 2 concentration levels, L4 and L8 (low and high, respectively), as well as for both QCM levels, in both media (Suppl. Tables 1, 2). Briefly, the classical approach uses normalized operational process specifications (OPSpec) charts, on which CV and bias are plotted on the x-axis and the y-axis, respectively, as percentages of TEa, to determine an operational point determining in turn candidate QC rules.[9,20] We investigated the sigma metric and the acceptable QC rules:

- at "low" TEa (slightly ≥TEo) and at arbitrarily chosen "high" TEa (see below),
- at high P_{ed} (90%) and arbitrarily chosen low P_{ed} (50%), and

Table 3. Reverse approach (Fig. 2B) for QC rule validation for commercial QCM (K9CON for serum, Liquicheck for urine), at low (insufficient), intermediate (optimized), and high (excessively high) TEa for the system, using EZ Rules 3, for $n = 2$ QCM levels.

Increasing TEa$_{QCM}$			Low TEa$_{QCM}$: 20%		Optimized TEa$_{QCM}$ ($\sigma = 5$)		High TEa$_{QCM}$: 50%	
Serum	QCM1	Sigma	$\sigma = 4.8^*$		TEa $= 20.8\%^*$		$\sigma = 12.1$	
	CV $= 4.1\%$	QC rule	P$_{ed}$	P$_{fr}$	P$_{ed}$	P$_{fr}$	P$_{ed}$	P$_{fr}$
	Bias $= 0.3\%$	1_{2S}	0.99	0.09	1	0.09	1.00	0.09
	TEo $= 8.5\%$	$1_{2.5S}$	**0.93**	**0.03**	**0.96**	**0.03**	**1.00**	**0.03**
		1_{3S}	0.80	0.00	0.86	0.00	**>0.98**	**0.00**
		$1_{3S}/2_{2S}/R_{4S}$	**0.91**	**0.01**	**0.94**	**0.01**	**1.00**	**0.01**
	QCM2	Sigma	$\sigma = 2.8$		TEa $= 35.5\%$		$\sigma = 7.1$	
	CV $= 7\%$	QC rule	P$_{ed}$	P$_{fr}$	P$_{ed}$	P$_{fr}$	P$_{ed}$	P$_{fr}$
	Bias $= 0.5\%$	1_{2S}	0.35	0.09	1	0.09	1.00	0.09
	TEo $= 14.5\%$	$1_{2.5S}$	0.16	0.03	**0.96**	**0.03**	**1.00**	**0.03**
		1_{3S}	0.06	0.00	0.86	0.00	>0.98	0.00
		$1_{3S}/2_{2S}/R_{4S}$	0.08	0.01	**0.94**	**0.01**	**1.00**	**0.01**
Urine	QCM1	Sigma	$\sigma = 3.5$		TEa $= 28.1\%$		$\sigma = 9.1$	
	CV $= 5.4\%$	QC rule	P$_{ed}$	P$_{fr}$	P$_{ed}$	P$_{fr}$	P$_{ed}$	P$_{fr}$
	Bias† $= 1.1\%$	1_{2S}	0.72	0.09	1	0.09	1.00	0.09
	TEo $= 11.9\%$	$1_{2.5S}$	0.48	0.03	**0.96**	**0.03**	**1.0**	**0.03**
		1_{3S}	0.25	0.00	0.86	0.00	>0.98	0.00
		$1_{3S}/2_{2S}/R_{4S}$	0.59	0.01	**0.94**	**0.01**	**1.0**	**0.01**
	QCM2	Sigma	$\sigma = 4.2$		TEa $= 23.6\%$		$\sigma = 10.9$	
	CV $= 4.5\%$	QC rule	P$_{ed}$	P$_{fr}$	P$_{ed}$	P$_{fr}$	P$_{ed}$	P$_{fr}$
	Bias† $= 1.1\%$	1_{2S}	0.91	0.09	1	0.09	1.00	0.09
	TEo $= 10.1\%$	$1_{2.5S}$	0.77	0.03	**0.96**	**0.03**	**1.0**	**0.03**
		1_{3S}	0.57	0.00	0.86	0.00	>0.98	0.00
		$1_{3S}/2_{2S}/R_{4S}$	0.69	0.01	**0.94**	**0.01**	**1.0**	**0.01**

Optimized TEa of the system, usually the intermediate one, has been set up for sigma approaching 5; one exception is QCM1 for serum, for which sigma already approaches 5 at "low" TEa (20%). The validated QC rules are those with **bold** P$_{ed}$ and P$_{fr}$, verifying at the same time P$_{ed} > 0.9$ and P$_{fr} < 0.05$. AB $=$ average bias; P$_{ed} =$ probability of error detection (by the QCM); P$_{fr} =$ probability of false rejection (by the QCM); QC $=$ quality control; QCM $=$ quality control material; RB $=$ range-based bias; $\sigma =$ sigma metric; SR $=$ spiking-recovery bias; TEo $=$ total allowable error; TEo $=$ observed total error; K9CON (Immulite Systems Control; Siemens); Liquicheck (Bio-Rad).

* For serum QCM1, performance parameters (CV, bias, TEo) were so good that the elected low TEa was already enough to generate a sigma approaching 5.

† For the urine QCM, given that no target values were provided by the manufacturer for this method, TE calculations have been done with the global bias of the comparison study (see discussion for justification).

- at $n = 2$ (2 levels analyzed once; results provided: all of the results of our study are for $n = 2$) or $n = 4$ (2 levels in duplicate; results not provided) measurements.

Low TEa was chosen at 20%, except for L4 for which it was set at 33%; high TEa was chosen at 50%. In urine, in rare occasions in which TEo was higher than these elected TEa, TEa was set as $=$ TEo.

Determination of the optimized TEa$_{QCM}$ and associated QC rules by the reverse approach of QC rule validation

Our "reverse approach" of QC rule validation (Fig. 2B) consisted of using EZ Rules 3 to generate the P$_{ed}$ of QC rules from the previously determined CVs and biases (P$_{fr}$ being fixed per rule):

- At low (insufficient), optimized ($\sigma = 5$), and high (excessive) TEa$_{QCM}$ for commercial QCM, to illustrate the impact of TEa$_{QCM}$ variation (or sigma variation) on QC rule efficiency (P$_{ed}$).
- At optimized TEa$_{QCM}$ only ($\sigma = 5$) for the spiked QCM L4 and L8, to compare with commercial QCMs.

We first summarized the biases and CVs, as well as the resulting TEo, for the concentration levels of interest (L4, L8, QCM1, QCM2) in serum and urine. For urine QCM, because there were no target values available, we considered the average bias only (see canine urine cortisol validation study[10]). We then investigated the resulting theoretical TEa$_{QCM}$ for defined sigma metrics, as well as the resulting P$_{ed}$ of the various QC rules (Fig. 2B) using EZ Rules 3. QC rules were accepted as candidates when both P$_{ed} > 90\%$ and P$_{fr} < 5\%$ were verified.

Table 4. Reverse approach (Fig. 2B) of QC rule validation for 2 cortisol concentrations L4 (38.6 nmol/L [1.4 µg/dL]) and L8 (552 nmol/L [20 µg/dL]), using the spiking-recovery bias, the between-run CV, and optimized TEa ($\sigma=5$) by EZ Rules 3.

Spiked sample QCM	QC rules	1_{2S}		$1_{2.5S}$		1_{3S}		$1_{3S}/2_{2S}/R_{4S}$	
	TE for $\sigma=5$	P_{ed}	P_{fr}	P_{ed}	P_{fr}	P_{ed}	P_{fr}	P_{ed}	P_{fr}
Serum cortisol									
L4	54%	1	0.09	0.96	0.03	0.86	0	0.94	0.01
L8	40%	1	0.09	0.96	0.03	0.86	0	0.94	0.01
Urine cortisol									
L4	90%	1	0.09	0.96	0.03	0.86	0	0.94	0.01
L8	43%	1	0.09	0.96	0.03	0.86	0	0.94	0.01

L4 = sample at a cortisol concentration of 38.6 nmol/L (1.4 µg/dL); L8 = sample at a cortisol concentration of 552 nmol/L (20 µg/dL); P_{ed} = probability of error detection (by the QCM); P_{fr} = probability of false rejection (by the QCM); QC = quality control; QCM = quality control material; σ = sigma metric; TE = total error.

Table 5. Summary of spiked samples (L4: 38.6 nmol/L = 1.4 µg/dL; L8: 552 nmol/L = 20 µg/dL) and QCM samples (QCM1 and QCM2; K9CON for serum, Liquicheck for urine) performance for cortisol measurement in dogs with the Immulite 2000 Xpi: between-run CV, bias (SR bias for spiked samples, target-value bias for serum commercial QCM, average bias from interlaboratory comparison-study for urine commercial QCM), and total error for sigma=2 (TEo) and sigma=5 (optimized TEa$_{QCM}$).

Level	Serum cortisol				Urine cortisol			
	CV(%)	Bias(%)	TEo ($\sigma=2$)	TEa$_{QCM}$ ($\sigma=5$)	CV(%)	Bias(%)	TEo ($\sigma=2$)	TEa$_{QCM}$ ($\sigma=5$)
L4	9.5	−6.50	25.6	54	9.6	42.3	61.4	90.3
L8	7.4	3.08	17.9	40	4.6	19.8	28.9	42.8
QCM1	4.1	−0.27	8.5	20.8	5.4	−1.09	11.9	28.1
QCM2	7	−0.52	14.5	35.5	4.5	−1.09	10.1	23.6

K9CON (Immulite Systems Control; Siemens); Liquicheck (Bio-Rad).

Results

For commercial QCM1 and QCM2

The reverse approach of QC rule validation for commercial QCMs (Table 3) yielded optimized TEa$_{QCM}$ ($\sigma=5$) of ~20% for QCM1 and ~35% for QCM2 for serum cortisol, and of ~28% for QCM1 and ~24% for QCM2 for urine cortisol. At fixed sigma ($\sigma=5$), P_{ed} was constant for a given QC rule (1_{2S}, $1_{2.5S}$, 1_{3S}, or $1_{3S}/2_{2S}/R_{4S}$) regardless of the QCM level or medium, illustrating the direct relationship between sigma (or TEa$_{QCM}$) and P_{ed}. At sigma=5, the acceptable QC rules were systematically $1_{2.5S}$ and $1_{3S}/2_{2S}/R_{4S}$, and the latter can be validated in this methodologic setting if the users are satisfied with the corresponding TEa$_{QCM}$.

To illustrate the influence of sigma or TEa$_{QCM}$ on P_{ed}, P_{ed} was also investigated at low TEa$_{QCM}$ (arbitrarily chosen at 20%) and high TEa$_{QCM}$ (arbitrarily chosen at 50%; Table 3). At low TEa$_{QCM}$, no QC rules were acceptable for any QCM (sigma was too low to generate acceptable P_{ed} according to our pre-established requirement of ≥0.90) except the serum QCM1, for which CV and bias were low enough to generate a sigma approaching 5 at low TEa$_{QCM}$. In other words, at low TEa$_{QCM}$, meaning low sigma, statistical QC cannot be relied on to achieve a high P_{ed} and low P_{fr} needed to effectively monitor the ongoing stable performance of the assays. At high TEa$_{QCM}$, all QC rules were acceptable for all QCMs (sigma was high enough to generate acceptable P_{ed}), except 1_{2S} that we disregarded initially because of subjectively excessive P_{fr}. Indeed, the P_{fr} of the immediate next QC rule ($1_{2.5S}$) is 0.03, making $1_{2.5S}$ much more desirable than 1_{2S}, decreasing by 3 times the risk of generating a false rejection.

For spiked sample QCMs: L4 & L8

The reverse approach of QC rule validation for spiked samples L4 and L8 considered as QCM was performed only at optimized TEa$_{QCM}$ ($\sigma=5$; Table 4). It yielded optimized TEa$_{QCM}$ ($\sigma=5$) of ~54% for L4 and ~40% for L8 for serum cortisol, and of ~90% for QCM1 and ~43% for QCM2 for urine cortisol. These results were very different from those with commercial QCMs.

Moreover, results (Tables 3, 4) provided a straightforward illustration that P_{ed} is fixed and constant per QC rule for a given sigma, and P_{fr} is fixed and constant for a given QC rule regardless of sigma. In other words, the same QC rules (1_{2S} and $1_{3S}/2_{2S}/R_{4S}$) were acceptable for commercial QCMs and spiked QCMs, but for very different associated TEa$_{QCM}$, illustrating the necessity of assessing QC rules based on the associated TEa$_{QCM}$ and P_{ed} as a whole.

Table 6. Summary of within-run CV(%) for spiked samples ($n=20$), between-run CV(%) for spiked samples ($n=20$ over 5 consecutive days), and QCM ($n=17$ for urine and $n=22$ for serum, over 1 mo).

	Within-run spiked sample	Between-run spiked sample	QCM sample
Serum low	1.4 μg/dL	1.4 μg/dL	K9CON QCM1: 7 μg/dL
CV	7.5%	9.5%	4.1%
Serum high	20 μg/dL	20 μg/dL	K9CON QCM2: 14 μg/dL
CV	4.7%	7.4%	7%
Urine low	1.4 μg/dL	1.4 μg/dL	Liquicheck QCM1: 6.7 μg/dL
CV	6.4%	9.6%	5.4%
Urine high	20 μg/dL	20 μg/dL	Liquicheck QCM2: 23.5 μg/dL
CV	2.5%	4.6%	4.5%

K9CON (Immulite Systems Control; Siemens); Liquicheck (Bio-Rad).

Comparison of commercial QCMs and spiked sample QCMs

The CV and bias (composing TEo) for commercial QCM samples and spiked samples were not equivalent (Table 5). In serum, the biases were much more marked in spiked samples, in which the bias was determined by spiking-recovery, than in commercial QCM samples, for which the bias is typically tailored. (This is why in urine, in the absence of commercial QCM available for the Immulite analyzers, we elected the AB.)

For between-run precision, commercial QCM CVs were also lower than spiked sample CVs. More specifically, when concentration levels of spiked samples and commercial QCM samples were close (higher concentrations levels), CVs were close; when the concentration level of spiked samples was much lower than the commercial QCM concentration (lower concentrations levels), spiked sample CVs were approximately twice as high as QCM CVs (Table 6). We concluded that the concentration level impacted the CV more than the actual nature of the QCM sample (commercial vs. spiked).

Illustration of relationships between concepts involved in statistical QC monitoring

Our multiple observations from the classical approach (Suppl. Tables 1, 2) and from the reverse approach (Tables 3, 4) of QC rule validation allowed us to identify and illustrate the complex relationships of the involved parameters (Fig. 4). Interactions between CV, bias, TEo, TEa, and σ are usually more easily acquired than interactions between TEa, σ, QC rules, P_{ed}, and P_{fr}.

When TEa increases:

- σ increases, as more σCV can be contained in TEa for a given bias.
- More QC rules can be accepted as candidates.
- P_{ed} improves (increases).
- P_{fr} is not impacted by the method performance or TEa, and is a defined function of QC rules and of the number of QCM levels (N, usually=2).
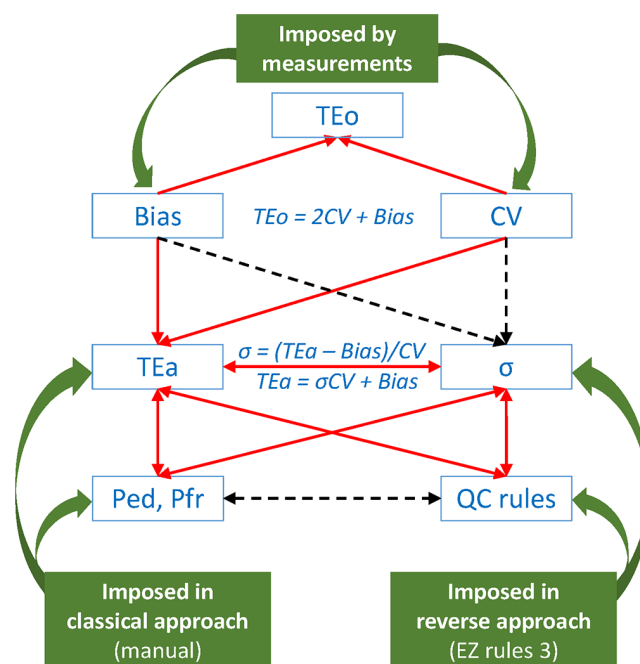

Figure 4. Influences of QC rule determinants on each other. Red plain single arrow: increases; red plain double arrow: vary in similar sense; black dotted single arrow: decreases; black dotted double arrow: vary in opposite sense, which means that when QC rules increase (from 1_{2S} to $1_{2.5S}$ to 1_{3S}), P_{ed} decreases (not desirable) and P_{fr} decreases (desirable). EZ Rules 3 QC design software; Ped=probability of error detection; Pfr=probability of false rejection; QC=quality control; σ=sigma metric; TEa=total allowable error; TEo=observed total error.

From the P_{ed} and P_{fr} perspective:

- Constant P_{ed} for a given QC rule at a given σ, regardless of the QCM.
- Increasing P_{ed} (improving) for a given QC rule with increasing σ, and increasing P_{ed} for a given σ with increasing QC rule stringency.
- Constant P_{fr} for a given QC rule, regardless of σ and regardless of the QCM.
- Decreasing P_{fr} (improving) with decreasing stringency of QC rules.

In other words, for a given sigma or a given TEa_{QCM}, P_{ed} increases (desirable) while P_{fr} increases (not desirable) with an increase of QC rule stringency (e.g., for simple QC rules: from 1_{3S} to $1_{2.5S}$ to 1_{2S}).

Discussion

In addition to proposing the validation of QC rules to monitor daily the testing of canine serum and urine cortisol in dogs with the Immulite 2000 Xpi, one of the main investigations of our work consisted of characterizing the major influence of both the nature of the QCM (commercial vs. spiked patient sample) and the analyte concentration of the QCM (imposed in commercial products vs. targeting IT in spiked samples) on the TEo (CV and bias) and on the candidate QC rules. The first and primary goal of our study was to establish a practical, achievable QC design. We used a reverse engineering approach to validate QC rules for QC monitoring with commercial QCMs. Indeed, our validation studies of canine serum[9] cortisol and canine urine[10] cortisol on the Immulite 2000 Xpi demonstrated the influence of the selected TEa_{QCM} on acceptable QC rules for QC daily monitoring, but did not allow straightforward identification of the most adequate QC rules. QC rules need to be stringent enough to allow satisfactory P_{ed} and P_{fr} of the system, but they also need to be simple and cost-effective to not waste money, time, and energy.[23] Because our previous validation studies were based on the classical approach of QC rule validation (Fig. 2A) using pre-selected TEa_{QCM} and observed performance parameters of bias(%) and CV(%) for $P_{ed} > 90\%$ and $P_{fr} \leq 5\%$,[22] we decided to reverse the approach (Fig. 2B) by imposing sigma and QC rules on the system instead, and to assess QC rules based on their resulting P_{ed}.

The reverse approach of QC rule validation is not a new validation method per se, but rather the application of QC rule validation from another perspective. It provides QC rules with associated P_{ed} & TEa_{QCM}, which laboratory users are entitled to accept or reject based on their own requirements. The reverse approach does not determine clinical TEa: the latter will have to be determined by broad expert consensus based on the needs for medical decision making. However, the ability to provide quality control for a particular amount of error should inform the clinical interpretation of the test, and clinician's desired clinical TEa may not always be technically achievable with the state-of-the-art instrument performance. The reverse approach requires users to have robust expertise in the field to elect a sigma metric with a resulting TEa_{QCM} mirroring as closely as possible the intended use of the generated patient results, meaning the TEa that a clinical expert may use for interpretation of the results. The intended use of the results is the key for the entire sigma metric theory, with or without the reverse approach. The generated QC rules are valid specifically to assure monitoring of the aforementioned level of quality for the intended use.

We believe that for the K9CON on the Immulite 2000 Xpi, $1_{2.5S}$ ($P_{ed} = 0.96$, $P_{fr} = 0.03$) for serum cortisol with TEa_{QCM} of ~20% and 35% (for QCM1 and QCM2, respectively) is satisfactory, given that we expect these levels of TEa_{QCM} to remain close to the clinically useful limit of a future consensus TEa, although it remains unclear if the future consensus TEa will extend as high as 35%. Similarly, we believe that for the Liquicheck on the Immulite 2000 Xpi, $1_{2.5S}$ ($P_{ed} = 0.96$ & $P_{fr} = 0.03$) for urine cortisol with TEa_{QCM} of ~27% and 24% (for QCM1 and QCM2, respectively) is satisfactory, given that we expect these levels of TEa_{QCM} to remain close to the clinically useful limit of a future consensus TEa. Indeed, if an assay cannot be controlled at a level that clinicians desire, this may provide impetus for manufacturers to improve the assay or clinicians to revise their opinions regarding clinical levels of TEa.

If the generated combination of QC rule, P_{ed}, and TEa_{QCM} appears satisfactory to laboratory users, then we suggest that the corresponding QC rule can be considered internally validated and used until a consensus TEa is determined in veterinary medicine for canine cortisol in serum and in urine. If the generated TEa_{QCM} appears too high to users, then the process could be repeated with a lower sigma and lower TEa_{QCM}, meaning with election of more stringent QC rules (or more QCM levels), or possibly with acceptation of a lower P_{ed}; however, we do not recommend decreasing P_{ed} to <0.90. For example, if the TEa_{QCM} of 35% for serum cortisol QCM2 is considered too high by some users, one could elect a sigma of 4, making a TEa_{QCM} of 28.5%, for which EZ Rules 3 would yield P_{ed} of 0.7 and P_{fr} of 0.03 for $1_{2.5S}$, or P_{ed} of 0.48 and P_{fr} of 0 for 1_{3S}. It is up to each laboratory to set their quality expectation until consensus TEa is determined. On the other hand, we do not believe that QC rules cannot be validated in the absence of consensus TEa, given that the latter would imply either complete ignorance of the used QC rule efficiency, or (even worse) useless QC monitoring of the concerned analytes, giving a false sense of security regarding monitoring of system stability. Non-statistical quality assurance practices, such as correlation with clinical signs, results of other evaluations, and results of serial testing, are also of particular importance with interpretation of endocrine results.

On commercial QCM, the TEa_{QCM} will need to allow at least a sigma >3 to generate acceptably simple QC rules for the system for routine testing for commercial purposes. If the sigma of the QCM is ~5, we recommend the use of QC rule $1_{2.5S}$. If the sigma of the QCM is between 3 and 5, more stringent QC rules may need to be elected to guarantee the lower level of TEa_{QCM}, and thus of clinical TEa. Because the acceptable QC rules are a direct function of the TEa_{QCM}, the TEa_{QCM} is directly conditioned by the clinical TEa, and a consensus about the clinical TEa has not yet been determined, we cannot prescriptively recommend the use of a particular QC rule. However, the $1_{2.5S}$ QC rule appears the most relevant one to use, keeping in mind the corresponding TE levels.

Our second goal was to illustrate the difference between state-of-the-art performance (precision, bias) of commercial QCMs (QCM1, QCM2) and spiked samples at relevant cortisol concentrations within real animal matrices (L4, L8). The use of QCM CV data should consider the QCM concentration. Given the similarity in CV when concentrations were similar and the differences in CV when concentrations were different, the difference in concentration is likely responsible for a large part of the difference in CV. On the other hand, the biases for both types of samples are viscerally different: for QCM samples, the bias quantifies a difference with a target tailored for the method (and is then expected to be low), whereas in spiked samples, the bias quantifies a difference of the method with an absolute target (and then may be expected to be significantly higher). The bias is then mostly responsible for the difference in TEo between commercial QCMs and spiked sample QCMs. The goal of QC monitoring is to guarantee the proper functioning of the system, not the absolute accuracy of the generated results, and in that regard, commercial QCM samples are sufficient. Using spiked samples introduces SR bias into the equation, which may be high for some methods, especially at low concentrations, yielding unrealistically low performance of the QC monitoring. Using the spiked samples as daily QCMs is not practical because it would either result in excessive TEa$_{QCM}$ or unreasonably stringent QC multirules, on top of being very inconvenient to generate and maintain. However, for interpretation of patient results, we strongly recommend the use of TEo from the samples at IT concentrations within the canine matrix, based on between-run precision and SR bias, for example in serum:

- L4 ± TEo = 38.6 nmol/L (1.4 µg/dL) ± 25.6% = 28.7–48.6 nmol/L (1.04–1.76 µg/dL)
- L8 ± TEo = 552 nmol/L (20 µg/dL) ± 17.9% = 452.5–651 nmol/L (16.4–23.6 µg/dL)

If TEo from commercial QCM were used instead, the resulting intervals would be 35.3–41.9 nmol/L and 472–632 nmol/L for L4 and L8, respectively, much narrower, and thus changing interpretations of results.

We pointed out that one of the main investigations of our work consisted of characterizing the major influence of the nature of the QCM and the analyte concentration of the QCM CV, bias, TEo, and candidate QC rules. In other words, we investigated if commercial QCMs and patient samples are commutable and interchangeable; this question is especially relevant, as often the QCM bias, and occasionally the QCM CV, are used in place of patient sample performances to determine the acceptability of a method. "Commutability" is defined as the "equivalence of the mathematical relationships between the results of different measurement procedures for a reference material and for representative samples from healthy and diseased individuals."[6,15] Therefore, if 2 types of samples (patient samples and QCM samples) are commutable for 2 measurement methods A and B, they give the same results when measured by method A and method B;

this is typically assessed by linear regression and prediction limits. Baral et al.[2] have pointed out that the original definition of commutability (equivalence of the mathematical relationships) did not only apply to accuracy (equivalent measurement results) but also encompassed precision (commutable material should also demonstrate equivalent CV). To facilitate the discussion about the latter points, Baral et al.[2] have used the term "interchangeability" when referring specifically to equivalence of precision between sample types. In their study[2] about biochemistry results on feline plasma pools versus commercial QCM, they determined that most analytes were commutable (same concentration results), but most analytes were not interchangeable (different CV). It is important to point out that commutability assessment compensates for the difference in analyte concentrations in the assessed samples (through linear regression), whereas interchangeability does not compensate for the difference in analyte concentrations when assessing the respective CV.

We have found in both serum and urine that the CV is greater in spiked patient samples (identical to true patient samples at common serum IT cortisol concentrations) than in commercial QCM samples, and then that patient samples and commercial QCM samples were not interchangeable; however, the concentration is, in our opinion, likely responsible for most of the differences. One study[14] has specifically investigated interchangeability between canine serum pools and commercial QCM for 3 analytes (urea, creatinine, C-reactive protein), and obtained different results depending on concentrations (below, within, or above RIs), stressing the importance of the concentration when assessing precision.

We have found in serum that the bias is much greater in spiked patient samples (identical to true patient samples at common serum IT cortisol concentrations) than in commercial QCM samples, and then that patient samples and commercial QCM samples were definitely not commutable: this is because target values for commercial QCM samples are tailored for these samples. As a consequence, the QCM bias may markedly underestimate the bias of the method. This is not problematic for daily QC monitoring, which is mostly focused on the precision of the method but may be problematic for method validation when intending to verify TEo < TEa, especially in veterinary endocrinology. The use of the QCM bias is especially appealing in biochemistry and hematology, given the high numbers of involved analytes and for convenience (a single analyzer allows investigation of precision and bias). The occasional use of the commercial QCM bias in the scientific literature to calculate the clinical TEo illustrates the link existing between QCM and clinical data.[5] This link is further illustrated by the use of ASVCP-recommended TEa for the purpose of QC rule validation.[5] The ASVCP guidelines provide TEa in biochemistry[8] and hematology[11] as quality goals: those are often envisaged as clinical quality goals (limiting the total error allowed for a test in order to maintain correct interpretation of results), but are also used as TEa$_{QCM}$ for QC rule validation. Thus, when a consensus recommendation for TEa has not been determined (as for cortisol in dogs),

it is relevant to document the theoretical TEa (clinical and QCM), which would be needed to afford a high sigma for the QCM (for example of 5), and thus the use of simple QC rules with high P_{ed} and low P_{fr}. The CV and bias used for TEo calculation should be analyzed closely for each study because authors may use different components (from QCM instead of from patient samples, or at concentrations markedly different from RI limits or ITs, etc.) than intuitively expected. In one study[5] of analytical performance of a dry chemistry analyzer, the authors computed TEo from the short-term precision (instead of between-run) and from the QCM bias (instead of the patient sample bias). Depending on analytes, as a result of conservation and stability issues, it is indeed not always possible to include between-run precision. Similarly, depending on analytes, it is not always possible to access a gold standard or even to create a spiked sample of determined concentration to assess the bias, hence QCM bias may occasionally be the only available bias for TE calculations. Determining the bias on commercial QCM instead of patient pools removes the matrix component of the equation,[20] especially relevant in veterinary medicine. We overcame the challenge of patient sample bias determination by using cortisol-free species-specific matrices spiked with cortisol at IT concentrations. To sum up, we emphasize that the TEo used to confirm adequacy of a method by confirming TEo<consensus TEa should use, if possible, bias and between-run CV from patient samples at relevant concentrations rather than QCM data; the QCM data should be reserved ideally for QC rule validation. This is especially relevant in endocrinology; in chemistry and hematology, using the QCM bias with patient sample CV to assess TEo<TEa for the method might be not ideal, but it is very convenient and economical.

Our third goal was to discuss the levels of TEa for cortisol in dogs based on state-of-the art performance of the method. Experts have stressed the critical importance of conducting research based on native samples, especially to characterize quality and performance associated with testing interpreted according to thresholds.[2] Echoing this recommendation, we document state-of-the-art performance of the method through proposed TEa and P_{ed} levels based on documented, relevant CV and bias, reinforcing model 3 for analytical performance specifications.[13] Once a consensus TEa is achieved by expert groups, such as the ASVCP/ESVCP, the American College of Veterinary Internal Medicine (ACVIM)/ECVIM, and the European Society of Veterinary Endocrinology (ESVE), this TEa goal can then be used to verify TEo<TEa as required for an acceptable method. We wish to stress that TEo<consensus TEa should be verified not only for QCM (which should be simple for commercial QCMs), but also for samples at relevant cortisol concentration (IT) within the correct animal matrix, such as L4 and L8 in our study. Given the levels of optimized TEa_{QCM} that we characterized for L4 and L8, it appears unlikely that future consensus TEa will be as high as sigma=5 for those; on the other hand, consensus TEa may reach sigma ≥5 for commercial QCMs. The Immulite 2000 Xpi method is unlikely to be discarded for canine serum

cortisol measurement; it is already widely used for clinical purposes. A consensus recommendation TEa for canine urine cortisol is, to date, unavailable; we suggest that the Immulite 2000 Xpi can be used with a documented, specific IT for urine cortisol:creatinine ratio, and knowledge of the corresponding TEo.

Our fourth goal was to use the example of cortisol in dogs to illustrate the complex relationships between the different concepts involved in statistical QC monitoring, to promote understanding within the veterinary community. There are multiple ways of entering this conceptual system, functioning in communication between what is imposed versus what consequently results. The CV and bias are always measured, and thus imposed within this system. Then, in the classical approach, TEa as well as P_{ed} and P_{fr} are imposed, and the resulting σ and QC rules are generated. In the reverse approach, σ and QC rules are imposed on this system, and the resulting TEa_{QCM} as well as the P_{ed} of the QC rules are generated. The reverse approach allows addition of more flexibility to the QC rule validation process, generating optimized TEa, with freely variable P_{ed} for the considered QC rules. This allows performance characterization and thus case-by-case internal validation of QC rules for analytes for which consensus TEa has not been determined. We find this model of interest for applying QC validation, determining aspects for improvement of existing assay methods, and in understanding the relationships of performance parameters, quality goals, and metrics for laboratory processes.

Finally, our study was performed on the former cortisol Siemens immunoassay, in which the anti-cortisol antibody changed late 2020. The new cortisol Siemens immunoassay contains an intrinsic correction formula compensating for the negative bias in veterinary samples (dogs, cats, horses) with the new antibody, and the kit is now labeled "Veterinary Cortisol" kit (VCO). The Immulite 2000 Xpi method will need to be reassessed with the new Siemens VCO for measurement of both serum and urine cortisol. Comparisons of canine serum cortisol values between the antibody in the kits we used and the new antibody yielded satisfactory results, characterized by a reasonably low negative bias.[7] On the other hand, for canine urine cortisol, there was a marked negative bias between the new and the former antibody, improved by the intrinsic correction factor, but not enough to allow the use of the same IT. In light of the consequent positive SR bias for cortisol in urine in our study, this may be an encouraging finding because the new antibody should in theory decrease, and thus improve, the positive SR bias for canine urine cortisol.[7]

Our observations on one Immulite 2000 Xpi will need to be repeated on several analyzers to verify the interlaboratory levels of TEa_{QCM}, and to determine if the future consensus about clinical TEa on canine serum and urine cortisol is higher than the most frequently observed TEa_{QCM}. A deciding factor for sorting out the TEa will be the level of performance that can be achieved routinely across multiple laboratories, and that we can control using these performance

parameters with a simple rule ($1_{2.5S}$ or 1_{3S}) and $n = 2$ (number of tested QCM levels).

ORCID iDs

Jeremie Korchia (iD) https://orcid.org/0000-0002-9344-6639
Kathleen P. Freeman (iD) https://orcid.org/0000-0003-1796-0158

Supplemental material

Supplemental material for this article is available online.

References

1. Arnold JE, et al. ASVCP guidelines: principles of quality assurance and standards for veterinary clinical pathology (version 3.0): developed by the American Society for Veterinary Clinical Pathology's (ASVCP) Quality Assurance and Laboratory Standards (QALS) Committee. Vet Clin Pathol 2019;48: 542–618.
2. Baral RM, et al. Commutability and interchangeability of commercial quality control materials with feline plasma for common biochemical analytes. Vet Clin Pathol 2016;45:300–310.
3. Behrend EN, et al. Diagnosis of spontaneous canine hyperadrenocorticism: 2012 ACVIM consensus statement (small animal). J Vet Intern Med 2013;27:1292–1304.
4. Feldman EC. Comparison of ACTH response and dexamethasone suppression as screening tests in canine hyperadrenocorticism. J Am Vet Med Assoc 1983;182:506–510.
5. Flatland B, et al. Analytical performance of a dry chemistry analyzer designed for in-clinic use. Vet Clin Pathol 2014;43: 206–217.
6. Franzini C, Ceriotti F. Impact of reference materials on accuracy in clinical chemistry. Clin Biochem 1998;31:449–457.
7. Graham P. Preliminary report on impact of Immulite 2000 cortisol antibody change October 2020 including evaluation of manufacturer recommended adjustment factors, 2020 Nov 3. https://www.esve.org/news/2020/20201109cortisolmeasurement_PreliminaryReport_Immulite2000Impact.pdf
8. Harr KE, et al. ASVCP guidelines: allowable total error guidelines for biochemistry. Vet Clin Pathol 2013;42:424–436.
9. Korchia J, Freeman KP. Validation study of canine serum cortisol measurement with the Immulite 2000 Xpi cortisol immunoassay. J Vet Diagn Invest 2021;33:844–863.
10. Korchia J, Freeman KP. Validation study of canine urine cortisol measurement with the Immulite 2000 Xpi cortisol immunoassay. J Vet Diagn Invest 2021;33:1052–1068.
11. Nabity MB, et al. ASVCP guidelines: allowable total error hematology. Vet Clin Pathol 2018;47:9–21.
12. O'Neil MJ, et al. The Merk Index. 15th ed. RSPC Publishing, 2013:4824.
13. Sandberg S, et al. Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2015;53:833–835.
14. Tecles F, et al. The number of replicates, and pooling versus individual measurements for analytical imprecision calculations: does it matter? Vet Clin Pathol 2020;49:112–118.
15. Vesper HW, et al. Reference materials and commutability. Clin Biochem Rev 2007;28:139–147.
16. Westgard J. CLIA requirements for analytical quality, 1992 Feb 28. https://www.westgard.com/clia.htm
17. Westgard JO. What is the purpose of a validation study? In: Westgaard JO, et al., eds. Basic Method Validation: Training in Analytical Quality Management for Healthcare Laboratories. 3rd ed. Westgard QC, 2008:27–36.
18. Westgard JO. How are the statistics calculated? In: Westgaard JO, et al., eds. Basic Method Validation: Training in Analytical Quality Management for Healthcare Laboratories. 3rd ed. Westgard QC, 2008:83–100.
19. Westgard JO. How can a manufacturer's claims be evaluated on the Sigma scale? In: Westgaard JO, et al., eds. Basic Method Validation and Verification: Training in Analytical Quality Management for Healthcare Laboratories. 4th ed. Westgard QC, 2016:135–152.
20. Westgard JO. How do you judge the performance of a method? In: Westgaard JO, et al., eds. Basic Method Validation: Training in Analytical Quality Management for Healthcare Laboratories. 3rd ed. Westgard QC, 2008:187–196.
21. Westgard JO. Analyzing and assessing risk. In: Westgard JO, ed. Six Sigma Risk Analysis: Designing Analytic QC Plans for the Medical Laboratory. Westgard QC, 2011:25–50.
22. Westgard JO. A safety net to catch analysis error. In: Westgard JO, ed. Six Sigma Risk Analysis: Designing Analytic QC Plans for the Medical Laboratory. Westgard QC, 2011:51–66.
23. Westgard JO. Selecting the right control materials. In: Westgard JO, ed. Basic Method Validation and Verification: Training in Analytical Quality Management for Healthcare Laboratories. 4th ed. Westgard QC, 2016:103–116.
24. Westgard JO. Selecting the right SQC procedure. In: Westgard JO, ed. Basic Method Validation and Verification: Training in Analytical Quality Management for Healthcare Laboratories. 4th ed. Westgard QC, 2016:135–152.
25. Westgard JO. Consolidated comparison of chemistry performance specifications, 2021 Sept 23. https://www.westgard.com/consolidated-goals-chemistry.htm